ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

INFORMATION TECHNOLOGY
DIVISION

Laboratory Research Computing

Gary Jung
IT Division

ITAC Meeting October 24, 2008

This Talk

- Introduce a new IT Lab-wide service
- Description of service
- Provide information on how to sign-up

Laboratory Research Computing

The LRC Project:

- $1M cluster system to make high performance computing more accessible to Berkeley Lab researchers

Drivers

- 38% of scientists depend on cluster computing for research.
- 69% of scientists are interested in cycles on a Lab-owned cluster.
  - Early-career scientists twice as likely to be 'very interested' than later-career peers

response rate for scientists: 37%
minimum confidence level: 95%

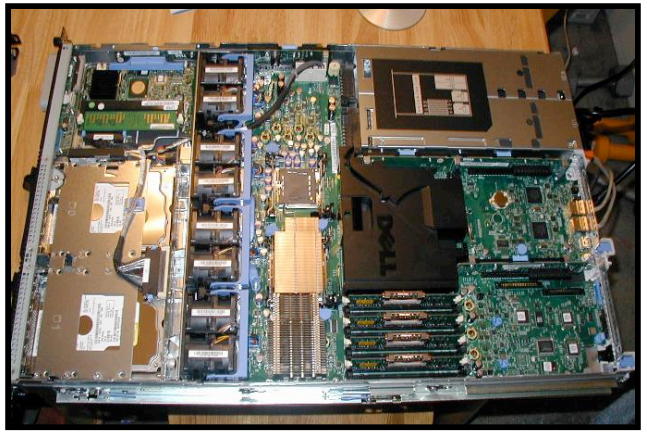
Goals

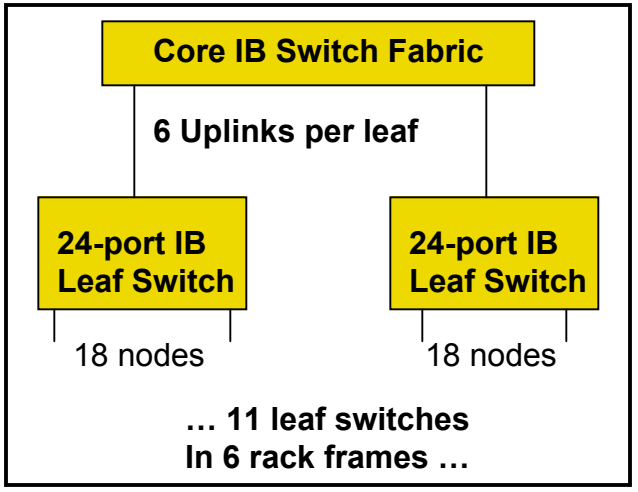
- Meet scientific need for compute cycles
- Provide onramp environment for scientists to benchmark applications in preparation for running on larger systems or applying for grants and supercomputing center allocations

Laboratory Research Computing

The Lawrencium Cluster

- Cluster Hardware
  - Dell PowerEdge 1950 III Linux cluster
    - ~ 198 dual-socket, quad-core processor compute nodes (1584 cores)
    - ~ 2.66 Ghz Intel Harpertown processors
    - ~ Dual channel 1333Mhz Front-side bus
    - ~ 16GB 667Mhz memory per compute node
    - ~ DDR Infiniband interconnect with 3:1 blocking (20Gb/s)
    - ~ 10 Interactive user nodes
- Performance
  - 12.6 Tflop/s Linpack (16.8 TF theoretical peak)

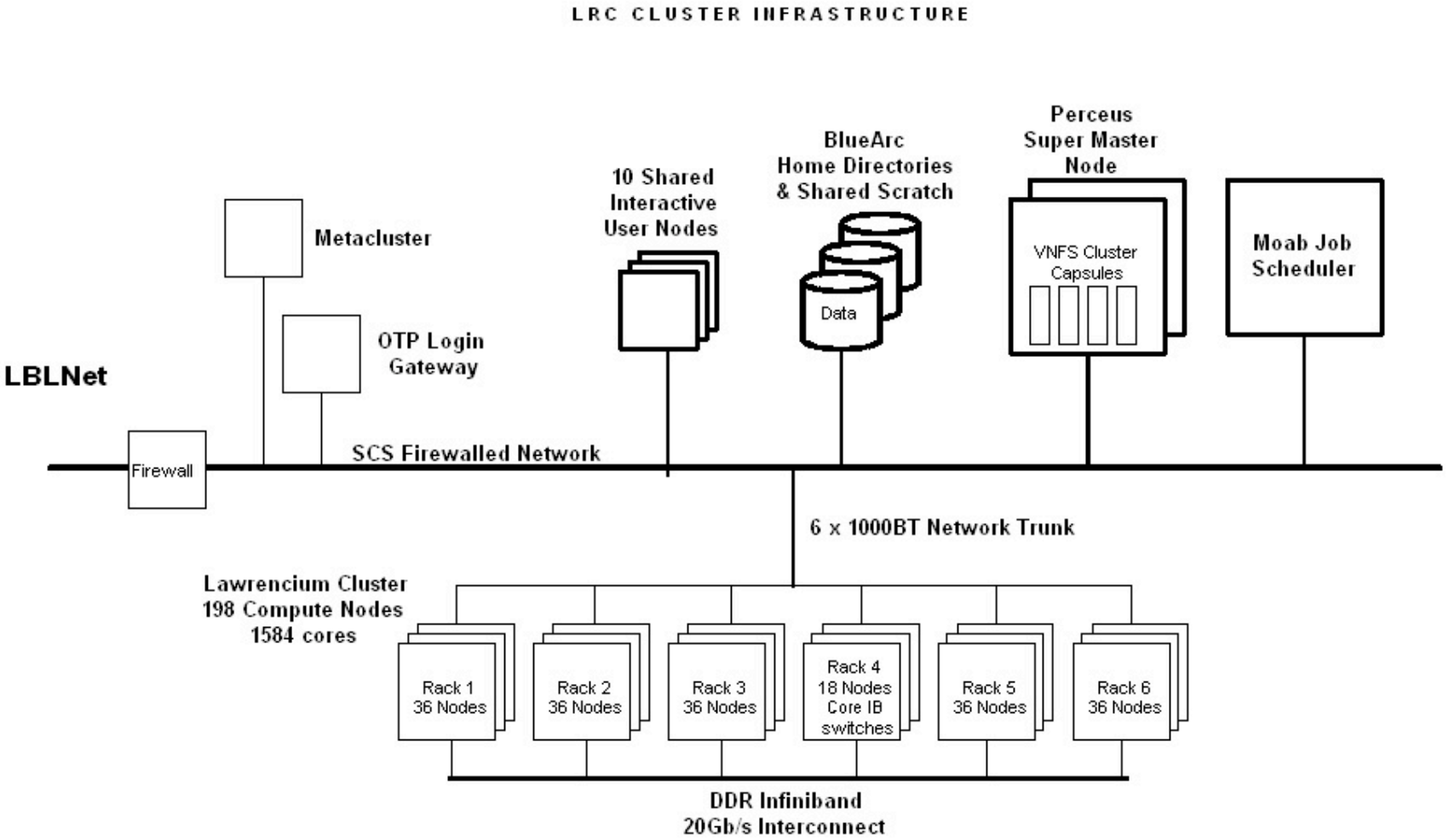




Lawrencium Storage

- Difficult to a find reasonably priced storage solution that can provide both reliability and support for parallel I/O.
- Solution: Two-tier storage strategy
- Use Bluearc high performance NFS storage subsystem for reliability and backups to get started
  - 19TB SATA storage for Home Directories
  - 13TB FC storage for shared scratch filesystem
- 45TB Lustre Parallel filesystem to be added later to handle parallel I/O needs
- Home filesystem will be provisioned across SCS-managed clusters to minimize the need to copy data between clusters.

LRC Configuration



Laboratory Research Computing

System Software

Centos 5.x operating system
Cluster Resources Moab job scheduler
Torque Resource manager
Gold Banking Software
OpenMPI 1.2.7 built with Intel compilers
Environment Modules

Compilers

Intel 10.1 Fortran and C++ compiler
GNU compilers 4.1.2
Portland Group Fortran compiler *

Debuggers

Etnus Totalview debugger (64 tokens)

Libraries

Intel Math Kernal libraries 10.0.4.023
(including BLAS & LAPACK)
FFTW (Fast Fourier Transform) 2.1.5 & 3.1.2
X11 libraries
GNU scientific libraries

Applications (To Date)

VASP* – Vienna Ab-initio Simulation Package
MEEP* – MIT Electromagnetic Equation Program
CCSM* – Community Climate System Model

Tools

HDF5* – Hierarchichal Data Format
NetCDF* – Network Common Data Format
NCO* – NetCDF Operators

Job Scheduling

Proposed Scheduling Policy

- Jobs are sent to the queue in a FIFO manner.
- Backfill is turned on. (which will cause some out-of-order execution).
- Maximum job limits on each execution queue
- Dedicated nodes (8 cores per node)

Execution Queue	Max # of nodes per job	Max # of jobs per user	Max # of total jobs running	Max time time per job
lr_debug	2	1	8	00:30:00
lr_small	8	8	No limit	72:00:00
lr_medium	32	2	4	24:00:00
lr_large	64	2	2	12:00:00

Allocations

Compute Cycles

- No formal allocations for the first 6 months (i.e from Dec 1, 2008 to May 31, 2008). Fairness will be handled by the scheduler.
- Every request for access is granted.
  - Need to be an LBNL employee.
  - Collaborations can be accommodated if the PI is LBNL
  - The use of allocations will be assessed after the 6 months
- Account request by filling an online form after PI registers project with SCS

Storage

- 10GB per user Home Directory storage
- No quotas on scratch at this time. Older files will be purged to manage space. This will be assessed later.
- Projects can purchase additional Bluearc storage to add to our servers to accommodate large data requirements.

Services

Staffing

- 1 FTE Cluster Administrator
- 1 FTE User Services consultant
  - General user help including how-to information
  - Accounts/allocations/job scheduling assistance
  - Porting and compiling assistance
  - Applications performance tuning

How to Get Started

Getting an Account

- Complete the short survey at http://LRC.lbl.gov/
- LBNL PIs should contact scs@lbl.gov to get their project setup for new account requests.
  - Send name and description of project along with survey
  - Usage (and possible future allocations) will be based by project
  - List of authorized users
- New account requests should go to the IT Help Desk starting in Dec
- Access will be exclusively via Cryptocard One-Time Password Tokens
- A nominal $25/mo. per user recharge will cover storage and backups of home directories

Project Timeline

May - Procurement bidding and award

June - Vendor factory build. Facilities preparation

Aug - System delivery and installation

Sep - Installation and software integration

Oct - Early access for debugging system

Nov - Preproduction use

Dec - General Availability

Feb - User Group Meeting

Current Status

Hardware burn-in completed. Developing user documentation

Early use in progress - 8 users (MSD, ESD, PHYS, CRD, AFRD, PBD) in various stages of getting their code ready

Initial reports on system and storage performance have been good

Early User Example

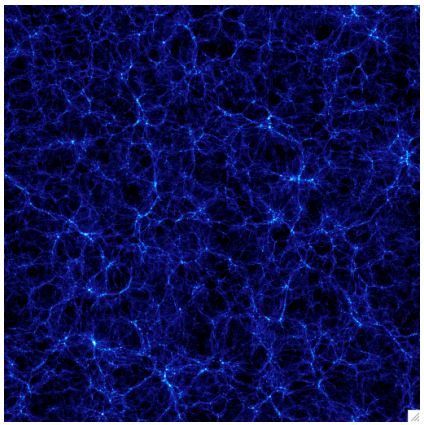
Science - Simulating the Formation of Structures in the Universe

PIs

- Nikhil Padmanabhan – Physics
- Martin White – Physics

Calculation

- The nature of the structures in the Universe provide important clues to the nature of dark matter and dark energy
- N-body simulations necessary to predict the effects of gravity in the nonlinear regime, and connect theory to observations



A slice through a cosmological N-body simulation showing the cosmic web of dark matter

Acknowledgements

Research Computing Task Force

Paul Adams (PBD, JBEI)
Paul Spellman (LSD)
William Miller (CSD)
David Schlegel (PHYS)
David Prendergast (MSD)
William Fawley (AFRD)
Craig Tull (NERSC)

LRC Allocations Committee

Paul Adams (PBD, JBEI)
Tom Daley (ESD)
Bill Fawley (AFRD)
David Lorenzetti (EETD)
William Miller (CSD)
Alan Poon (NSD)
David Prendergast (MSD)
Dan Pulsifer (ENG)
Paul Spellman (LSD)
Nobumichi Tamura (ALS)
Francesca Verdier (NERSC)